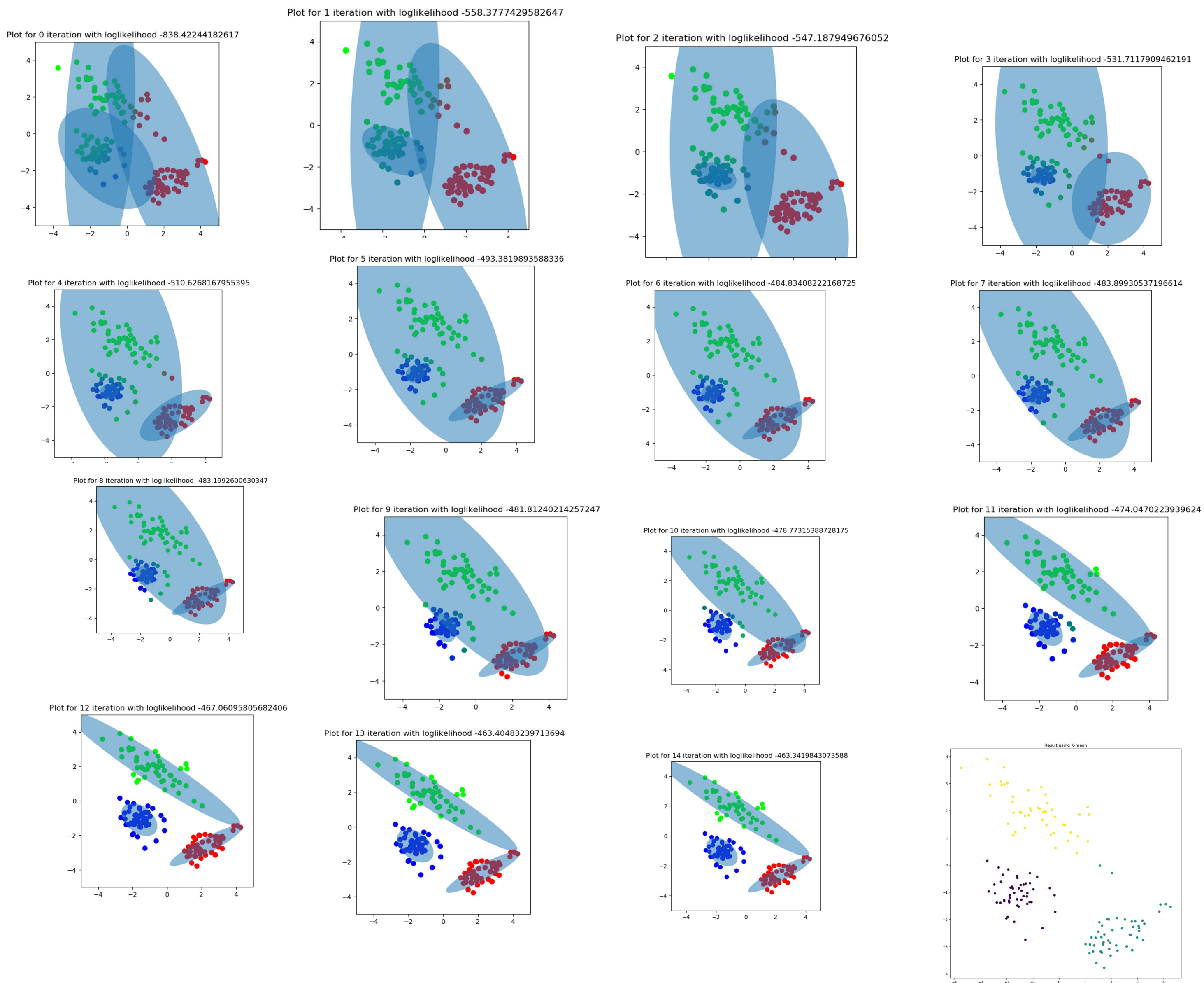
Thursday, April 2, 2020 6:33 PM



By comparison, the result using K-mean and Em are roughly the same except two intermediate points.

2. Solution

(a) $H(p,q) = -\sum_i p_i \cdot \log_2 q_i$

$= -\left(\frac{1}{8}\cdot\log_2\frac{1}{4} + \frac{1}{2}\cdot\log_2\frac{1}{8} + \frac{1}{8}\cdot\log_2\frac{1}{8} + \frac{1}{8}\cdot\log_2\frac{1}{4} + \frac{1}{8}\cdot\log_2\frac{1}{4}\right)$

$= \frac{1}{4} + \frac{3}{2} + \frac{3}{8} + \frac{2}{8} + \frac{2}{8}$

$= \frac{21}{8}$

$H(q,p) = -\sum_i q_i \cdot \log_2 p_i = \frac{11}{4} \neq H(p,q)$

(b) $H(p) = -\sum_i p_i \cdot \log p_i$

$= -\left(\frac{1}{8}\cdot\log\frac{1}{8} + \frac{1}{2}\cdot\log\frac{1}{2} + 3\cdot\frac{1}{8}\cdot\log\frac{1}{8}\right)$

$= \frac{3}{8} + \frac{1}{2} + \frac{9}{8}$

$= 2$

$H(q) = -\sum_i q_i \cdot \log_2 q_i$

$= -\left(\frac{1}{4}\cdot\log_2\frac{1}{4} + 2\cdot\frac{1}{8}\cdot\log_2\frac{1}{8} + 2\cdot\frac{1}{4}\cdot\log_2\frac{1}{4}\right)$

$= \frac{2}{4} + \frac{6}{8} + \frac{4}{4}$

$= \frac{9}{4}$

(c) $D_{KL}(p|q) = H(p,q) - H(p,p)$

$= \frac{21}{8} - 2$

$= \frac{5}{8}$

$\blacksquare$

3. Solution

(a) $$X^T X = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

Let $A = X^T X$, next find the eigenvector of A.

$$|A - \lambda I| = \begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = (2-\lambda)^2 - 1 = 0 \Rightarrow 2-\lambda = \pm 1$$

$$\Rightarrow \lambda = 1 \text{ or } \lambda = 3$$

For $\lambda_1 = 1$,

$$(A - I)\vec{v}_1 = \vec{0} \Rightarrow \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \vec{v}_1 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$$

For $\lambda_2 = 3$,

$$(A - 3I)\vec{v}_2 = 0 \Rightarrow \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \vec{v}_2 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

Hence

$$V = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

$$AA^T = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

Let $B = AA^T$, next find the eigenvector of B.

$$|B-\lambda I| = 0 \Rightarrow \begin{vmatrix} 2-\lambda & 1 & 1 \\ 1 & 1-\lambda & 0 \\ 1 & 0 & 1-\lambda \end{vmatrix} = 0$$

$$\Rightarrow (2-\lambda)(1-\lambda)^2 - (1-\lambda) - (1-\lambda) = 0$$

$$\Rightarrow (1-\lambda)\left[(2-\lambda)(1-\lambda)-2\right] = 0$$

$$\Rightarrow (1-\lambda)(\lambda^2-3\lambda) = 0 \Rightarrow \lambda_1 = 0,\ \lambda_2 = 3,\ \lambda_3 = 1$$

For $\lambda_1 = 0$,

$$B\vec{v}_1 = \vec{0} \Rightarrow \vec{v}_1 = \begin{pmatrix} -1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix}$$

For $\lambda_2 = 3$,

$$(B-3I)\vec{v}_2 = \vec{0} \Rightarrow \begin{pmatrix} -1 & 1 & 1 \\ 1 & -2 & 0 \\ 1 & 0 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \vec{0} \Rightarrow \vec{v}_2 = 1/\sqrt{6}\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$$

For $\lambda_3 = 1$,

$$(B-I)\vec{v}_3 = \vec{0} \Rightarrow \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \vec{0} \Rightarrow \vec{v}_3 = 1/\sqrt{2}\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$

Hence

$$U = \begin{pmatrix} 2/\sqrt{6} & 0 & -1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \end{pmatrix}$$

$$X = \underbrace{\begin{pmatrix} 2/\sqrt{6} & 0 & -1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \end{pmatrix}}_{U} \underbrace{\begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}}_{\Sigma} \underbrace{\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}}_{V^T} ,$$

(b) Note that

$$XV = \begin{bmatrix} | & & | & & | \\ \vec{u}_1 & \cdots & \vec{u}_r & \cdots & \vec{u}_m \\ | & & | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}$$

$$= \begin{bmatrix} | & | & & | \\ \sigma_1 \vec{u}_1 & \sigma_2 \vec{u}_2 & \cdots & \sigma_r \vec{u}_r \\ | & | & & | \end{bmatrix}$$

For any pairs of column $i \neq j$, the dot product is

$$(\sigma_i \vec{u}_i) \cdot (\sigma_j \vec{u}_j) = \sigma_i \sigma_j (\vec{u}_i \cdot \vec{u}_j)$$

$\vec{u}$'s are orthonormal

$$= \sigma_i \sigma_j (0)$$

$$= 0$$

Hence the columns of T are pairwise orthogonal.

4. Solution

```
import numpy as np
from sklearn import decomposition
from sklearn import datasets



X = datasets.load_diabetes().data
row, col = X.shape
for column_ix in range(col):
   X[:, column_ix] -= np.mean(X[:, column_ix])
   X[:, column_ix] /= np.std(X[:, column_ix])

pca = decomposition.PCA()
pca.fit(X)
print("The V matrix is \n", np.matrix.transpose(pca.components_))
print("The singular values are \n ", pca.singular_values_)
print("The 3 most important components for the first 10 data-points is \n", pca.transform(X)[:10, :3])
```

```
The V matrix is
 [[ 0.21643101  0.04437151  0.49466811 -0.4140095  -0.68686389 -0.2258505
  -0.10953821  0.01493468  0.00810057  0.00326309]
 [ 0.18696711 -0.38654811 -0.10685833 -0.67986052  0.37345612  0.04173103
  -0.06760551  0.44293966 -0.00210552  0.00366069]
 [ 0.3031625  -0.15628061  0.1675317   0.49982533  0.12935936 -0.4031419
  -0.51985787  0.39294187  0.04237751  0.00824809]
 [ 0.2717397  -0.13825564  0.51356804 -0.01966734  0.48689014 -0.27276274
   0.32064908 -0.47736435  0.0271941  -0.00322111]
 [ 0.34325493  0.57302669 -0.0685867  -0.06839533  0.12917415  0.00540864
   0.07364908  0.12941351 -0.04203984  0.70977447]
 [ 0.35186062  0.45593985 -0.26969438 -0.16777384  0.11673143 -0.1332572
  -0.23054011 -0.19131121 -0.35931549 -0.56319605]
 [-0.28243639  0.50624287  0.38602787 -0.07602005  0.24499115  0.1063716
  -0.00753445  0.32463641  0.48124771 -0.31744413]
 [ 0.42883325 -0.06818423 -0.38068121  0.0079212  -0.14364377 -0.0339454
   0.07123619 -0.18058834  0.77381656 -0.09059464]
 [ 0.37861731 -0.0261893   0.0636315   0.26442742 -0.1516611   0.17873005
   0.64731345  0.44966002 -0.18945947 -0.26446735]
 [ 0.32218282 -0.0849466   0.27684271  0.08708624  0.03138792  0.80506447
  -0.35727279 -0.1666087  -0.01527381  0.0026109 ]]
The singular values are
  [42.17466853 25.68276971 23.08755816 20.55043949 17.10806903 16.32182255
 15.39999097 13.84514267  5.88365535  1.94518745]
The 3 most important components for the first 10 data-points is
 [[ 0.58720767 -1.94682793  0.58923299]
 [-2.83161209  1.37208454  0.02791506]
 [ 0.27214757 -1.63489803  0.73927034]
 [ 0.04931005  0.38225333 -2.01303697]
 [-0.75645071  0.81196754 -0.05725853]
 [-3.96635524 -0.38105927 -0.33738317]
 [-1.99378667 -0.80553831 -0.71219915]
 [ 2.07586704  1.82792114  0.52492352]
 [ 0.60303259 -0.88125266 -0.07671973]
 [-0.21215262 -0.49290431 -0.81436321]]
```

5. Solution

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + a_t \qquad \text{— (1)}$$

(a) Taking expectation on (1),

$$\mathbb{E}[r_t] = \phi_0 + \phi_1 \cdot \mathbb{E}[r_{t-1}] + \phi_2 \mathbb{E}[r_{t-2}].$$

By stationary, $\mathbb{E}[r_t] = \mathbb{E}[r_{t-1}] = \mathbb{E}[r_{t-2}] = \mu$, thus we have

$$\mu = \phi_0 + \phi_1 \mu + \phi_2 \mu$$

$$\Rightarrow (1 - \phi_1 - \phi_2)\mu = \phi_0$$

$$\Rightarrow \mu = \frac{\phi_0}{1 - \phi_1 - \phi_2}$$

Hence, $\mathbb{E}[r_t] = \dfrac{\phi_0}{1 - \phi_1 - \phi_2}$

(b) Using $\phi_0 = (1 - \phi_1 - \phi_2)\mu$, the AR(2) model can be written as

$$r_t - \mu = \phi_1 (r_{t-1} - \mu) + \phi_2 (r_{t-2} - \mu) + a_t$$

Multiply both sides by $r_{t-s} - \mu$ and take expectation

$$\mathbb{E}[(r_t - \mu)(r_{t-s} - \mu)] = \phi_1 \mathbb{E}[(r_{t-1} - \mu)(r_{t-s} - \mu)] + \phi_2 \mathbb{E}[(r_{t-2} - \mu)(r_{t-s} - \mu)]$$

which implies

$$Cov(r_t, r_{t-s}) = \phi_1 Cov(r_{t-1}, r_{t-s}) + \phi_2 Cov(r_{t-2}, r_{t-s}) \quad -(2)$$

and thus

$$\gamma(s) = \phi_1 \gamma(s-1) + \phi_2 \gamma(s-2)$$

Divide both sides by $\gamma(0)$,

$$\Rightarrow \quad \rho(s) = \phi_1 \rho(s-1) + \phi_2 \rho(s-2), \quad \forall s \geq 2$$

Substitute $s=1$

$$\rho(1) = \phi_1 \rho(0) + \phi_2 \rho(-1)$$

Since $\rho(0) = 1$ and $\rho(-1) = \rho(1)$

$$\rho(1) = \phi_1 + \phi_2 \rho(1)$$

$$\Rightarrow \rho(1) = \frac{\phi_1}{1-\phi_2},$$

Hence

$$\rho(s) = \phi_1 \rho(s-1) + \phi_2 \rho(s-2) \text{ for } \forall s \geq 2$$

and $\rho(1) = \dfrac{\phi_1}{1-\phi_2}$

$\blacksquare$